Supporting Information

Pathway Analysis for Drug Repositioning Based on Public Database Mining

Yongmei Pan, Tiejun Cheng, Yanli Wang*, Stephen H. Bryant*

National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, 8600 Rockville Pike, Bethesda, MD 20894, United States

1. The whole process of the data-mining and analysis for the pathway analysis regarding celecoxib

Step 1: to find the PubChem cid of celecoxib:
http://www.ncbi.nlm.nih.gov/pccompound/?term=celecoxib

Step 2: retrieval of drug targets from MMDB, PubChem and GEO
MMDB:
http://www.ncbi.nlm.nih.gov/structure?linkname=pccompound_structure&from_uid=2662
PubChem: http://pubchem.ncbi.nlm.nih.gov/assay/assay.cgi?cid=2662
GEO: http://www.ncbi.nlm.nih.gov/gds/?term=celecoxib

Step 3: retrieval of Gene ids of drug targets in the Gene database. Use drug target PTGS2 as an example.
http://www.ncbi.nlm.nih.gov/gene/?term=PTGS2

Step 4: pathway retrieval based on the obtained drug targets. All pathways involving each obtained drug target regarding each drug were identified by large-scale access to the Biosystems database using E-utilities with Ruby script (attached as Supporting Information). Following is an example of how to retrieve pathways involving a drug target "PTGS2" by using its Gene id 5743 via accessing Biosystem database.
http://www.ncbi.nlm.nih.gov/biosystems/?term=5743[geneid]+AND+pathway[type]+AND+Homo[Organism]

Step 5: find pathways common to the primary and secondary targets. The pathways were ranked in terms of the numbers of retrieved drug targets involved in each pathway.

Step 6: retrieve titles of each obtained pathway. Following is an example of how to retrieve a pathway by using a Biosystems ID (BSID) obtained from step 4. The significance of each pathway involving drug MOAs and clinical functions were examined. The Ruby scripts for large-scale access to Gene and Biosystems database and the afterward data analysis (involving steps 3-6) are attached as Supporting Information.

http://www.ncbi.nlm.nih.gov/biosystems/?term=198912

2. Detail discussion of relevant pathways identified for each drug (except gefinitib, celecoxib, tamoxifen, and tretinoin that are discussed in detail in the main text)

Adenosine is an antiarrhythmic agent for treatment of heart disorders [1]. The crystal structures of the primary target ADORA2A (PDB: 2YDO) and six secondary targets in complex with adenosine (Supporting Information, Table S3) were retrieved via the short-cut link to MMDB at the PuBChem Compound database. Six more protein targets were retrieved from BioAssay database. The primary target (ADORA2A/2B) shared biological pathway of Vascular Smooth Muscle Contraction (BSID: 96530) with secondary targets including ADCY2, PRKACA and PRKACB, indicating that adenosine can regulate vascular contraction by activation of adenosine receptors, and supporting its role as an antiarrhythmic agent for treatment of heart disorders [1]. A few pathways regarding GRCP signaling transduction were also obtained (Supporting Information, Table S4), possibly because there were retrieved proteins involved in the cAMP pathway (such as ADCY2 and PRKACA/B), a step regulated by the GPCR signal transduction [2].

Doxorubicin kills cancer cells by intercalating DNA, causing failure of topoisomerase II and DNA replication [3, 4]. It is used to treat multiple cancers, such as bladder, breast, stomach, lung, kidney, nerve, and blood cancers as well as leukemia and Hodgkin's lymphoma, and is also used for certain patients with liver, prostate, endometrium, and pancreas cancers[5]. No structure of human protein in complex with doxorubicin was available. There were nine proteins and twenty eight regulated genes identified via BioAssay and GEO DataSets. The retrieved targets led to a few cancer-related pathways (Table 2), including those regarding general and particular cancers, such as pancreatic, breast, lung, prostate, endometrial, colorectal, bladder, and gastric cancers, as well as acute myeloid leukemia, melanoma, and glioma. The involvement of the target genes and proteins into various cancer-related pathways supported the broad-spectrum usage of doxorubicin in cancer therapy. There were also relevant pathways about apoptosis that accounts for the drug effects of doxorubicin, such as the pathways of Apoptosis (BSID: 198797) and Apoptosis Modulation and Signaling (BSID: 198822).

Similar to doxorubicin, actinomycin is a chemotherapy drug that kills tumor cells by binding to DNA in transcription complex and therefore preventing RNA elongation [6]. Fourteen regulated genes and two proteins were obtained by checking GEO and BioAssay databases. Due to its extreme cytotoxicity, the clinical use of actinomycin is limited [7] and it is mainly used for rare cancers such as gestational trophoblastic neoplasia, Wilms' tumor, rhabdomyosarcoma, Ewing's sarcoma, and malignant

hydatidiform mole [8]. Actinomycin witnessed similar pathways to doxorubicin, including those responsible for apoptosis and various cancers except melanoma, glioma, bladder and gastric cancers that were indicated by pathways retrieved for doxorubicin. The reasons that the two drugs shared similar pathways might be that: 1) they both have their drug effects by interacting with DNA; 2) the affected genes were obtained from the same publication [9] which yielded overlapping genes between them. Most of obtained cancer pathways pointed to cancers other than those currently treated by actinomycin (Table 2). The involvement of actinomycin into the cancer pathways is reasonable, considering its effects on DNA transcription and apoptosis, and is consistent with previous studies showing effectiveness of actinomycin on apoptosis and various cancer cells [10-15]. Different from the four drugs with relevant pathways that might indicate information of drug repurposing, the relevant pathways retrieved by actinomycin may not indicate its new clinical uses, considering its high cytotoxicity [7] and the resistance to the agent induced by cancel cells [10].

Mifepristone is a synthetic steroid used as an abortifacient drug by targeting the progesterone receptor (PGR) [16, 17]. The structure of glucocorticoid receptor (NR3C1) bound with mifepristone (PDB: 1NHZ) was retrieved from MMDB. The primary target PGR and 14 more protein targets and 433 genes responsible for all kinds of biological pathways [18] were retrieved by checking BioAssay and GEO DataSets databases. Three biological pathways pertinent to its clinical effects were identified: the Oocyte meiosis pathway (BSID: 126909) shared by PGR,BUB1, CDK2, CHP, MAD2L1,YWHAG, the Progesterone-mediated oocyte maturation pathway (119304) shared by PGR, BUB1, CDK2, HSP90AB1, and MAD2L1, and the pathway of Ovarian Infertility Genes (BSID: 198801) shared by PGR, ATM, and VDR. Although only around ten of the more than 400 genes were involved in these pathways, they were the only identified pathways that were responsible for a biological process in a cell (Supporting Information, Table S4). Other pathways were general pathways regarding gene expression, signal transduction, and nuclear receptors, possibly due to the broad range of retrieved genes responsible for all kinds of cellular functions.

Rosiglitazone is an antidiabetic drug by binding to peroxisome proliferator-activated receptor (PPAR), therefore increasing insulin sensitivity in cells [19]. Besides the primary target PPAR, four and fourteen affected proteins and genes were characterized by checking BioAssay and GEO databases (Table 1), contributing to two pathways that supported the primary target protein PPAR and its clinical function: the Insulin Signaling pathway (BSID: 83090 shared by PPAR, IRS2, PCK1, and PDE3B) and the PPAR signaling pathway (BSID: 83042 shared by PPAR, CPT1B, FABP4, FABP7, and PCK1). The above pathways both accounted for the MOAs of drug effects of rosiglitazone.

Fluorouracil is an anti-cancer drug by irreversible inhibition of thymidylate synthase (TYMS), therefore blocking the biosynthesis of DNA and causing cell cycle arrest and apoptosis [20]. It is mainly used for treatment of colorectal and pancreatic cancer [21]. The crystal structure of uridine phosphorylase 1 (UPP1) bound with fluorouracil (PDB: 3NBQ) was obtained. Eighty genes and five interacting proteins (including primary target TYMS) were identified by checking GEO DataSets and BioAssay databases. Pathways accounting for the MOA of fluorouracil were retrieved, such as: 1) Pyrimidine Metabolism (BSID: 82946, shared by TYMS, UPP1, RRM2B, DUT, TK1, and RRM2); 2) Pyrimidine Deoxyribonucleotides de novo Biosynthesis I (BSID: 782380 shared by TYMS, RRM2B, DUT, and RRM2); 3) Pyrimidine Metabolism (BSID: 106281 shared by TYMS, UPP1, DUT, and TK1). A few pathways about cell cycle and mitosis were identified, including 1) Cell Cycle (BSID: 530733 shared by TYMS and 20 regulated genes); 2) Cell Cycle, Mitotic (BSID: 105765 shared by TYMS and 17 genes); 3) Mitotic G1-G1/S phases (BSID: 160941); 4) G1/S Transition (BSID: 105769) (Table 1). The four pathways are consistent with the clinical effect of fluorouracil that is attributed to cell cycle arrest and apoptosis. Another pathway was the Integrated Pancreatic Cancer Pathway (BSID: 711360) shared by TYMS and 16 identified genes, supporting the usage of fluorouracil for pancreatic cancer.

Vincristine is a mitotic inhibitor that is used as a chemotherapy agent [22]. By binding to tubulin, vincristine disrupts microtubule structures and therefore causing mitosis failure in cells. Two and eleven target proteins and genes were retrieved from BioAssay and via GEO DataSets databases. Two pathways accounting for cell mitosis were identified: 1) the Cell Cycle (BSID: 530733) shared by tubulin, MYC, MDM2, and GMNN; 2) Cell Cycle, Mitotic (BSID: 105765) shared by tubulin, MYC, and GMNN. The above pathways are responsible for the biological functions of vincristine in a cell and therefore accounting for MOA behind its drug effects.

Bortezomib is a proteasome inhibitor for treatment of relapsed multiple myeloma and mantle cell lymphoma [23, 24]. Twenty eight affected genes and eleven protein targets were retrieved by checking GEO and BioAssay databases. Two pathways underlying the MOA of bortezomib were identified: 1) the Protein Processing in Endoplasmic Reticulum (BSID: 167325 shared among proteasome, CAPN1, HSPA1A/B, CHL1, and BCL2); 2) the Parkin-Ubiquitin Proteasomal System pathway (BSID: 700638 shared by proteasome, SNCA, CUL1, and HSPA1A/B).

Propofol is a hypnotic agent by targeting γ-aminobutyric acid (GABA) A receptor [25, 26]. Table 1 demonstrates that eleven proteins and eighty eight genes were retrieved via BioAssay and GEO databases. Most identified pathways were about the transmission of neurontransmittors in CNS (Supporting

Information, Table S4). Only one of them, or the GABAergic synapse pathway (BSID: 377263) was directly pertinent to the MOA underlying its clinical uses. This pathway is shared by a GABA A receptor subunit (GABRA1-A6/B1-B3/G1-G3) and glutaminase (GLS), one of key enzymes that are responsible for GABA synthesis in brain [27, 28]. The lack of relevant pathways might be due to the limited pathways the genes participated (including the PGC-alpha, G-CSF_survival, Fatty acid oxidation, and DNA damage singaling pathways [29]) that are not directly responsible for the GABA signaling.

Valproic acid (VPA) is a drug for treatment of seizure disorders and unstable mood conditions by mainly affecting neurontransmiter levels via targeting GABA transaminase (ABAT), an enzyme that breaks down GABA [30]. There were 196 genes and 3 proteins (including primary protein ABAT) retrieved via checking GEO DataSets and BioAssay. Neuronal System pathway (BSID: 106513) was identified among ABAT, GNAI1, GNG11, and ALDH5A1; GABAergic Synapse pathway (BSID: 377263) was common among ABAT, GNAI1, and GNG11, where GNAI1 and GNG11 are G protein subunits that pass down signals from GABAB receptor in the GABAergic pathway [31]. The identified genes regulated a variety of cellular processes [32, 33]. It's surprising that the above pathways only comprised 4 or 3 of them. Considering there are only 89 genes displayed in the GABAergic synapse pathway, most of retrieved genes might be the down-stream proteins regulated by GABA-G protein pathway and are not included as involved proteins in this pathway. Only two protein targets (excluding ABAT) were available at the BioAssay database. Identification of more secondary target proteins might be needed so that a GABA-related pathway can be identified among G proteins and other proteins than the primary protein ABAT.

Two drugs, disulfiram and triiodothyronine failed to identify relevant pathways to its clinical functions. Triiodothyronine (known as T3) is a ligand of thyroid hormone receptors (THRs) that affects all physiological processes in the body [34]. It can be used for the treatment of depressive disorders [35, 36] and as a supplement for fat loss [37, 38]. A few crystal structures, including those of THRs (THRA and THRB), PCNA and AR in complex with T3 were obtained from MMDB (Table S3). There were eleven more protein targets and 275 genes retrieved for triiodothyronine. Only general pathways regarding gene expression and nuclear receptors were identified. The failure to find relevant pathway for triiodothyronine might be due to: 1) the retrieval of nuclear receptors in BioAssay leading to the identified general pathways regarding gene expression; 2) the retrieved genes responsible for variety of cellular functions that were not directly related to the pathways pertinent to its primary target protein thyroid receptors. The involvement of this drug in almost every physiological process in the body may also account for the lack of identified particular pathways to a cellular function.

Disulfiram is mainly used for treatment of chronic alcoholism by inhibiting acetaldehyde dehydrogenase (ALDH2 and ALDH1A1 in Table 1), an enzyme that plays an important role in conversion of alcohol into harmless acetic acid [39]. Fifty nine proteins including the primary target protein were retrieved from BioAssay, while no genes obtained by checking GEO DataSets. A few pathways regarding molecular metabolism and oxidation were retrieved (Supporting Information, Table S4), but no pathway specific to alcohol mechanism was obtained, except for the Ethanol oxidation pathway (BSID: 105715) shared by ALDH2 and ALDH1A1. Considering only 10 genes are involved in this pathway, the retrieved target proteins might be up- or down- stream proteins of this pathway.

References

1. Mitchell, J.; Lazarenko, G., Wide QRS complex tachycardia. Diagnosis: Supraventricular tachycardia with aberrant conduction; intravenous (IV) adenosine. *CJEM* **2008,** 10, (6), 572-3, 581.
2. Gilman, A. G., G proteins: transducers of receptor-generated signals. *Annu Rev Biochem* **1987,** 56, 615-49.
3. Fornari, F. A.; Randolph, J. K.; Yalowich, J. C.; Ritke, M. K.; Gewirtz, D. A., Interference by doxorubicin with DNA unwinding in MCF-7 breast tumor cells. *Mol Pharmacol* **1994,** 45, (4), 649-56.
4. Momparler, R. L.; Karon, M.; Siegel, S. E.; Avila, F., Effect of adriamycin on DNA, RNA, and protein synthesis in cell-free systems and intact cells. *Cancer Res* **1976,** 36, (8), 2891-5.
5. CLINIC, M., Doxorubicin (Intravenous Route) In Micromedex: 2012.
6. Sobell, H. M., Actinomycin and DNA transcription. *Proc Natl Acad Sci U S A* **1985,** 82, (16), 5328-31.
7. Takusagawa, F.; Carlson, R. G.; Weaver, R. F., Anti-leukemia selectivity in actinomycin analogues. *Bioorg Med Chem* **2001,** 9, (3), 719-25.
8. Initiative, T. S. H. C., Actinomycin-D. In 2013.
9. Lehnhardt, M.; Klein-Hitpass, L.; Kuhnen, C.; Homann, H. H.; Daigeler, A.; Steinau, H. U.; Roehrs, S.; Schnoor, L.; Steinstraesser, L.; Mueller, O., Response rate of fibrosarcoma cells to cytotoxic drugs on the expression level correlates to the therapeutic response rate of fibrosarcomas and is mediated by regulation of apoptotic pathways. *BMC Cancer* **2005,** 5, 74.
10. Ng, C. P.; Zisman, A.; Bonavida, B., Synergy is achieved by complementation with Apo2L/TRAIL and actinomycin D in Apo2L/TRAIL-mediated apoptosis of prostate cancer cells: role of XIAP in resistance. *Prostate* **2002,** 53, (4), 286-99.
11. Kleeff, J.; Kornmann, M.; Sawhney, H.; Korc, M., Actinomycin D induces apoptosis and inhibits growth of pancreatic cancer cells. *Int J Cancer* **2000,** 86, (3), 399-407.
12. Sella, A.; Aggarwal, B. B.; Kilbourn, R. G.; Bui, C. A.; Zukiwski, A. A.; Logothetis, C. J., Phase I study of tumor necrosis factor plus actinomycin D in patients with androgen-independent prostate cancer. *Cancer Biother* **1995,** 10, (3), 225-35.
13. Tsuruga, M.; Dang, Y.; Shiono, Y.; Oka, S.; Yamazaki, Y., Differential effects of vitamin E and three hydrophilic antioxidants on the actinomycin D-induced and colcemid-accelerated apoptosis in human leukemia CMK-7 cell line. *Mol Cell Biochem* **2003,** 250, (1-2), 131-7.
14. Moore, D. H.; Blessing, J. A.; Dunton, C.; Buller, R. E.; Reid, G. C., Dactinomycin in the treatment of recurrent or persistent endometrial carcinoma: A Phase II study of the Gynecologic Oncology Group. *Gynecol Oncol* **1999,** 75, (3), 473-5.

15. Guo, L.; Fan, L.; Ren, J.; Pang, Z.; Ren, Y.; Li, J.; Wen, Z.; Qian, Y.; Zhang, L.; Ma, H.; Jiang, X., Combination of TRAIL and actinomycin D liposomes enhances antitumor effect in non-small cell lung cancer. *Int J Nanomedicine* **2012,** 7, 1449-60.
16. Fiala, C.; Gemzel-Danielsson, K., Review of medical abortion using mifepristone in combination with a prostaglandin analogue. *Contraception* **2006,** 74, (1), 66-86.
17. Heikinheimo, O.; Kekkonen, R.; Lahteenmaki, P., The pharmacokinetics of mifepristone in humans reveal insights into differential mechanisms of antiprogestin action. *Contraception* **2003,** 68, (6), 421-6.
18. Yin, P.; Roqueiro, D.; Huang, L.; Owen, J. K.; Xie, A.; Navarro, A.; Monsivais, D.; Coon, J. S. t.; Kim, J. J.; Dai, Y.; Bulun, S. E., Genome-wide progesterone receptor binding: cell type-specific and shared mechanisms in T47D breast cancer cells and primary leiomyoma cells. *PLoS One* **2012,** 7, (1), e29021.
19. Mayerson, A. B.; Hundal, R. S.; Dufour, S.; Lebon, V.; Befroy, D.; Cline, G. W.; Enocksson, S.; Inzucchi, S. E.; Shulman, G. I.; Petersen, K. F., The effects of rosiglitazone on insulin sensitivity, lipolysis, and hepatic and skeletal muscle triglyceride content in patients with type 2 diabetes. *Diabetes* **2002,** 51, (3), 797-802.
20. Longley, D. B.; Harkin, D. P.; Johnston, P. G., 5-fluorouracil: mechanisms of action and clinical strategies. *Nat Rev Cancer* **2003,** 3, (5), 330-8.
21. CLINIC, M., Fluorouracil (Intravenous Route, Injection Route) In 2011.
22. Johnson, I. S.; Armstrong, J. G.; Gorman, M.; Burnett, J. P., Jr., The Vinca Alkaloids: A New Class of Oncolytic Agents. *Cancer Res* **1963,** 23, 1390-427.
23. Gelman, J. S.; Sironi, J.; Berezniuk, I.; Dasgupta, S.; Castro, L. M.; Gozzo, F. C.; Ferro, E. S.; Fricker, L. D., Alterations of the intracellular peptidome in response to the proteasome inhibitor bortezomib. *PLoS One* **2013,** 8, (1), e53263.
24. Bonvini, P.; Zorzi, E.; Basso, G.; Rosolen, A., Bortezomib-mediated 26S proteasome inhibition causes cell-cycle arrest and induces apoptosis in CD-30+ anaplastic large cell lymphoma. *Leukemia* **2007,** 21, (4), 838-42.
25. Kotani, Y.; Shimazawa, M.; Yoshimura, S.; Iwama, T.; Hara, H., The experimental and clinical pharmacology of propofol, an anesthetic agent with neuroprotective properties. *CNS Neurosci Ther* **2008,** 14, (2), 95-106.
26. Trapani, G.; Altomare, C.; Liso, G.; Sanna, E.; Biggio, G., Propofol in anesthesia. Mechanism of action, structure-activity relationships, and drug delivery. *Curr Med Chem* **2000,** 7, (2), 249-71.
27. Petroff, O. A., GABA and glutamate in the human brain. *Neuroscientist* **2002,** 8, (6), 562-73.
28. Schousboe, A.; Waagepetersen, H. S., GABA: homeostatic and pharmacological aspects. *Prog Brain Res* **2007,** 160, 9-19.
29. Lucchinetti, E.; Hofer, C.; Bestmann, L.; Hersberger, M.; Feng, J.; Zhu, M.; Furrer, L.; Schaub, M. C.; Tavakoli, R.; Genoni, M.; Zollinger, A.; Zaugg, M., Gene regulatory control of myocardial energy metabolism predicts postoperative cardiac function in patients undergoing off-pump coronary artery bypass graft surgery: inhalational versus intravenous anesthetics. *Anesthesiology* **2007,** 106, (3), 444-57.
30. Rosenberg, G., The mechanisms of action of valproate in neuropsychiatric disorders: can we see the forest for the trees? *Cell Mol Life Sci* **2007,** 64, (16), 2090-103.
31. Chen, K.; Li, H. Z.; Ye, N.; Zhang, J.; Wang, J. J., Role of GABAB receptors in GABA and baclofen-induced inhibition of adult rat cerebellar interpositus nucleus neurons in vitro. *Brain Res Bull* **2005,** 67, (4), 310-8.
32. Stamatopoulos, B.; Meuleman, N.; De Bruyn, C.; Mineur, P.; Martiat, P.; Bron, D.; Lagneaux, L., Antileukemic activity of valproic acid in chronic lymphocytic leukemia B cells defined by microarray analysis. *Leukemia* **2009,** 23, (12), 2281-9.

33. Wood, J. R.; Nelson-Degrave, V. L.; Jansen, E.; McAllister, J. M.; Mosselman, S.; Strauss, J. F., 3rd, Valproate-induced alterations in human theca cell gene expression: clues to the association between valproate use and metabolic side effects. *Physiol Genomics* **2005,** 20, (3), 233-43.
34. Triiodothyronine. In *Mosby's Medical Dictionary*, Elsevier: 2009.
35. Kelly, T.; Lieberman, D. Z., The use of triiodothyronine as an augmentation agent in treatment-resistant bipolar II and bipolar disorder NOS. *J Affect Disord* **2009,** 116, (3), 222-6.
36. Kelly, T. F.; Lieberman, D. Z., Long term augmentation with T3 in refractory major depression. *J Affect Disord* **2009,** 115, (1-2), 230-3.
37. Lanni, A.; Moreno, M.; Lombardi, A.; de Lange, P.; Silvestri, E.; Ragni, M.; Farina, P.; Baccari, G. C.; Fallahi, P.; Antonelli, A.; Goglia, F., 3,5-diiodo-L-thyronine powerfully reduces adiposity in rats by increasing the burning of fats. *FASEB J* **2005,** 19, (11), 1552-4.
38. Lombardi, A.; Lanni, A.; Moreno, M.; Brand, M. D.; Goglia, F., Effect of 3,5-di-iodo-L-thyronine on the mitochondrial energy-transduction apparatus. *Biochem J* **1998,** 330 (Pt 1), 521-6.
39. Wright, C.; Moore, R. D., Disulfiram treatment of alcoholism. *Am J Med* **1990,** 88, (6), 647-55.